Fundamental matrix estimation for binocular vision measuring system used in wild field

Yan Nian[12] and Wang Xiang-jun[12] and Liu Feng[12]

[1] State Key Laboratory of Precision Measuring Technology and Instruments, Tianjin University, Tianjin, 300072, China;
[2] MOEMS Education Ministry Key Laboratory, Tianjin University, Tianjin, 300072, China

ABSTRACT

A method has been proposed to estimate the fundamental matrix of a positing and monitoring binocular vision system with a long working distance and a large field of view. Because of the long working distance and large field of view, images grabbed by this system are seriously blurred, leading to a lack of local features. The edge points are acquired using the Canny algorithm firstly, then the pre-matched points are obtained by the GMM-based points sets registration algorithm, and eventually the fundamental matrix are estimated using the RANSAC algorithm. In actual application, two cameras are 2km away from the object, the fundamental matrix are figured out, and the distance between each point and the corresponding epipolar line is less than 0.8 pixel. Repeated experiments indicate that the average distances between the points and the corresponding epipolar lines are all within 0.3 pixel and the deviations of the distances are all within 0.3 pixel too. This method takes full advantage of the edges in the environment and does not need extra control points, whats more, it can work well in low SNR images.

Keywords: Canny, GMM, Fundamental Matrix, RANSAC

1. INTRODUCTION

The fundamental matrix is the basic object that represents the geometric information between two views in the pinhole camera model.[1]

Obtaining three dimensional information from images taken from different viewpoints is a vital and challenging task in computer vision. However, as the measured datas in images are just pixel coordinates, there are only two ways[2] that can be used to perform this task:

The first one is to establish a model which relates image coordinate system to world coordinate system, and to estimate all the parameters corresponding to such a model. This model can be acquired by camera calibration, which typically computes the 3×4 projection matrix $\hat{\mathbf{P}}$. The 11 parameters of this projection matrix account for both intrinsic and extrinsic camera calibration parameters. However it is not always possible to assume that cameras can be calibrated accurately, in some cases, it is even impossible to calibrate cameras.

The second approach is to use projection information, which needs no intrinsic parameters of the cameras. This approach requires only image coordinate system information which relates to the same world coordinate system from different viewpoints, thus a much more smaller number of parameters have to be estimated.

If we only have to estimate the fundamental matrix and do not care about the exact intrinsic and extrinsic parameters of the camera, the second method of course has more advantages. It is much easier to accomplish and needs to estimate less parameters, and there is no need to get information of the world coordinate system. it does benefit us a lot especially under wild-field condition. It is not easy to get accurate world coordinate information in the wild field. Besides, it is also a great challenge to correspond one point in world coordinate system to one in image coordinate system. Both reasons make it difficult to acquire accurate calibration information.

Further author information: (Send correspondence to YanNian)
Yan Nian: E-mail: yannian89@163.com, Telephone: +81 02227403395
Wang Xiang-jun: E-mail: xdocuxjw@vip.163.com
Liu Feng: E-mail: tjuliufeng@tju.edu.cn

International Symposium on Optoelectronic Technology and Application 2014: Image Processing and Pattern Recognition, edited by Gaurav Sharma, Fugen Zhou, Proc. of SPIE Vol. 9301, 93010S · © 2014 SPIE
CCC code: 0277-786X/14/$18 · doi: 10.1117/12.2070319

Moreover, it takes too much time to obtain these information, which should be taken into consideration in practical applications.

For the second method, we should get sufficient corresponding points in two images. After that, we can use different kinds of methods to estimate the fundamental matrix. All these methods can be divided into two categories: the linear algorithm and the nonlinear algorithm. The linear algorithm can estimate fundamental matrix in linear time, but because it ignores the constraint between parameters and the error criterion lacks physical meanings, its precision is low. This kind of algorithm typically includes Linear Least-square method, singular value decomposition, etc. The nonlinear algorithm maintains a higher precision, because it takes the physical meanings into account. The biggest difference between different iterative algorithms is the choice of different error criteria,[3] such as algebraic distance, symmetric epipolar distance or Sampson distance.

In practical applications, because of the outliers and the error of the detected points, robust algorithms[4] are used to eliminate the outliers and reduce the error, like RANSAC, LMedS, M-estimation and MLESAC.

All these algorithms require us to get enough corresponding points. In wild-field case, we can get the corresponding points through two methods. The first one is to add extra control points and get them in two images, which, as is mentioned above, is not practical. The second method is to use the image processing method to extract feature points. This method is a prevalent access to get enough corresponding points. The state-of-art algorithms, like Harris corner,[5] SIFT,[6] ASIFT,[7] MSER,[8] can do quite well in many cases. But in wild-field case, the long working distance makes the images quite blurred and we cannot easily extract enough local features accurately.

Inspired by contour points match algorithm[9] and points sets registration algorithm[10111213],[14] we propose an algorithm to cope with the problem of fundamental matrix estimation for binocular vision system used in wild field. Firstly, we acquire the edge points using canny edge detector, and then we get the pre-matched points by the GMM based point set registration algorithm, at last, we get the fundamental matrix using the RANSAC algorithm. This method takes full advantage of the edge information in the environment and does not need extra control points. It can work well in low SNR images as well.

The remainder of this paper is organized as follows, the theory of the GMM based point set registration is introduced in Section (2.1). Section (2.2) presents the theory of fundamental matrix estimation using RANSAC. Section (3) describes the experiments and compares our algorithm with ASIFT. Finally, the research is concluded in Section (4).

2. THEORY OF FUNDAMENTAL MATRIX ESTIMATION

2.1 GMM Based Point Set Registration

First, we obtain the edge of the monitored area using canny edge detector.[15] We denote the edge point set in the left image by $\mathbf{X}_{N\times 3} = (\mathbf{x}_1, \cdots, \mathbf{x}_N)^T$, and denote the edge point set in the right image by $\mathbf{Y}_{M\times 3} = (\mathbf{y}_1, \cdots, \mathbf{y}_M)^T$, where $\mathbf{x}_i, (i = 1 \cdots N)$ and $\mathbf{y}_i, (i = 1 \cdots M)$ are homogeneous coordinates of edge points. N and M indicate the number of the points in each set respectively.

The probability density function of a general Gaussian mixture is defined[13] as

$$p(\mathbf{x}|\mu_i, \Sigma_i) = \sum_{i=1}^{k} \omega_i \phi(\mathbf{x}|\mu_i, \Sigma_i) \tag{1}$$

where

$$\phi(\mathbf{x}|\mu_i, \Sigma_i) = \frac{1}{2\pi\sqrt{|\Sigma_i|}} \exp[-\frac{1}{2}(\mathbf{x} - \mu_i)^T \Sigma_i^{-1} (\mathbf{x} - \mu_i)] \tag{2}$$

We make three assumptions :

1. The edge point set $\mathbf{Y}$ is the GMM centroids and the edge point set $\mathbf{X}$ is the point set generated by GMM;
2. The number of Gaussian components is the number of the points in $\mathbf{Y}$, that is M;

3. All components of the GMM share a equal membership probability $P(m) = \frac{1}{M}$ and a equal isotropic covariances σ^2.

Then the GMM probability density function is

$$p(\mathbf{x}) = \sum_{m=1}^{M} P(m)p(\mathbf{x}|m) \tag{3}$$

where $p(\mathbf{x}|m) = \frac{1}{2\pi\sigma^2} e^{-\frac{\|\mathbf{x}-\mathbf{y}_m\|}{2\sigma^2}}$. In order to cope with the noise and outliers, we add an additional uniform distribution[14] $p(\mathbf{x}|M+1) = \frac{1}{N}$ to the gaussian mixture model. Denoting the weight of the uniform distribution as ω, $0 \leq \omega \leq 1$, the mixture model takes the form

$$p(\mathbf{x}) = \omega \frac{1}{N} + (1-\omega) \sum_{m=1}^{M} \frac{1}{M} p(\mathbf{x}|m) \tag{4}$$

According to theory of maximum likelihood, we estimate a set of θ by maximizing the likelihood, or by minimizing the negative log-likelihood function. θ contains all the transportation information between $\mathbf{X}$ and $\mathbf{Y}$.

$$E(\theta, \sigma^2) = -\sum_{n=1}^{N} log \sum_{m=1}^{M+1} P(m)p(\mathbf{x}_n|m) \tag{5}$$

In fact, we cannot directly minimize the Equation (5) by computing its derivative or through other methods. We choose the EM algorithm to estimate θ and σ^2. The EM algorithm contains two steps, the E-Step and the M-Step.

E-Step Use the Bayes' theorem to compute a posteriori probability distributions $P^{\text{old}}(m|\mathbf{x}_n)$ of mixture components according to Equation (7)

$$P(m|\mathbf{x}_n) = \frac{P(m)p(\mathbf{x}_n|m)}{p(\mathbf{x})} \tag{6}$$

$$= \frac{\exp[\frac{1}{2}\|\frac{\mathbf{x}_n - \mathcal{T}(\mathbf{y}_m,\theta)}{\sigma}\|]}{\sum_{k=1}^{M} \exp[\frac{1}{2}\|\frac{\mathbf{x}_n - \mathcal{T}(\mathbf{y}_m,\theta)}{\sigma}\|] + 2\pi\sigma^2 \frac{\omega}{1-\omega}\frac{M}{N}} \tag{7}$$

M-Step Estimate new parameter values by minimizing the expectation of the function in Equation (8)

$$Q = -\sum_{n=1}^{N}\sum_{m=1}^{M+1} P^{\text{old}}(m|\mathbf{x}_n) log(P^{\text{new}}(m)p^{\text{new}}(\mathbf{x}_n|m)) \tag{8}$$

we can rewrite Equation (8) into Equation (9).

$$Q(\theta, \sigma^2) = \frac{1}{2\sigma^2} \sum_{n=1}^{N}\sum_{m=1}^{M} P^{\text{old}}(m|\mathbf{x}_n)\|\mathbf{x}_n - \mathcal{T}(\mathbf{y}_m, \theta)\|^2 + N_p log\sigma^2 \tag{9}$$

where $N_p = \sum_{n=1}^{N}\sum_{m=1}^{M} P^{\text{old}}(m|\mathbf{x}_n)$

Repeat the E-Step and M-Step until it converges.

We use $\mathcal{T}(\mathbf{Y}, v) = \mathbf{Y} + v(\mathbf{Y})$ to define the transformation[14] for the nonrigid point set registration process of EM optimization .

2.2 Fundamental Matrix Estimation With RANSAC

Epipolar geometry describes the constraints satisfied by two perspective projections $\mathbf{x} = (u, v, 1)$ and $\mathbf{x}' = (u', v', 1)$ of the same physical 3D point on two different images from different view points. For pinhole cameras, the relation can be written in the following form

$$\mathbf{x}'^{T} F \mathbf{x} = 0 \tag{10}$$

where F, the fundamental matrix, is a 3×3 matrix with rank 2 that depends on the rigid motion between the two image planes and the camera parameters. We can interpret the Equation (10) like this: the point $\mathbf{x}'$ of the right image plane is on the epipolar line Fm whose line equation $au' + bv' + c = 0$ is obtained from $\mathbf{x} = (u, v, 1)$ by $(a, b, c)^T = F(u, v, 1)$, and vise verse.

With enough(at least eight) corresponding points $\mathbf{x}_i \leftrightarrow \mathbf{x}_i'$, the Equation (10) can be used to compute the unknown fundamental matrix F.

$$(u_i' u_i, v_i u_i', u_i', u_i v_i', v_i v_i', v_i', u_i, v_i, 1)\mathbf{f} = 0 \tag{11}$$

where $\mathbf{f} = (F_{11}, F_{12}, F_{13}, F_{21}, F_{22}, F_{23}, F_{31}, F_{32}, F_{33})^T$

Assume that we have n corresponding points, then we can obtain a system of linear equations with n equations from the Equation (11)

$$\mathbf{A}\mathbf{f} = \mathbf{0} \tag{12}$$

where

$$\mathbf{A} = \begin{bmatrix} u_1' u_1 & v_1 u_1' & u_1' & u_1 v_1' & v_1 v_1' & v_1' & u_1 & v_1 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ u_n' u_n & v_n u_n' & u_n' & u_n v_n' & v_n v_n' & v_n' & u_n & v_n & 1 \end{bmatrix}$$

The case $n = 8$ corresponds to the eight-points algorithm.[16] The solution of $f = 0$ is unacceptable. Moreover, the rank of F is 2, which has to be imposed afterward because of numerical stability of f computation. A common way to handle this problem is to compute a SVD of A and get a singular vector associated to the smallest singular value, this process is equivalent to the least square formulation, and then set the smallest singular value of F to zero.

In section 2.1, we use the GMM based point set registration algorithm to match the edge point set roughly. So we cannot use these corresponding points directly. We choose the RANSAC algorithm to eliminate the outliers and estimate the fundamental matrix.

3. EXPERIMENT

In this section, we present the results obtained under real working condition. Two cameras are both Basler cameras with lens whose focal lengths are $24mm$. The camera model is piA2400-17gc/gm, and the camera has a resolution of 2448×2050 and pixel size of $3.45 \times 3.45\mu m$. In experiment the cameras are both about 2km away from monitored objects.

First, we normalize[17] the coordinates of the edge points obtained by canny edge detector, because the large range of the coordinates makes the estimation of the fundamental matrix quite sensitive to noise. We implement this by shifting the center of the image coordinates to the center of the point set and scaling the point set to an average distance of $\sqrt{2}$. Then we use the GMM based algorithm to accomplish the point set registration. The result is shown in Fig. (1).

At last, we use RANSAC algorithm to find out inliers and estimate the fundamental matrix. And the matched points are shown in Fig. (2). In Fig. (3), we show the result of the ASIFT algorithm.[7] Our algorithm can find more correct matches than ASIFT. Although ASIFT find 14 matches, there are several points quite close to each other, it will make the fundamental matrix estimation unstable. And our algorithm can not only find matches in artificial objects but also in hill edges, it can be used in other environment where there are no artificial objects. After getting the fundamental matrix, we can calculate distance between one point and its corresponding epipolar

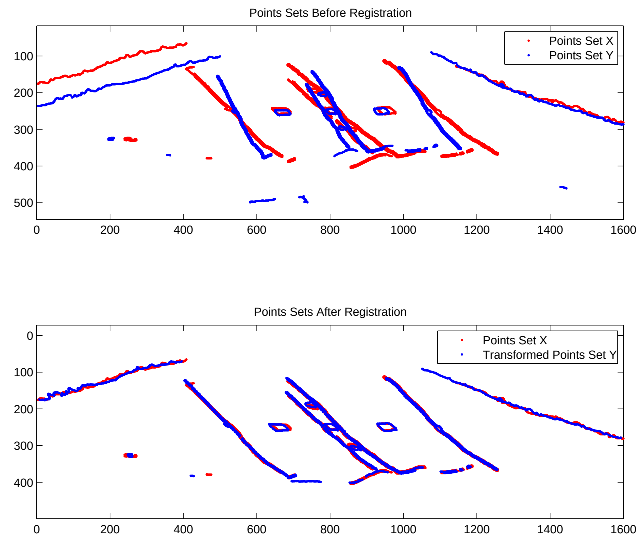

Figure 1. Top: The edge point set $\mathbf{X}$ and $\mathbf{Y}$ obtained by Canny edge detector. Bottom: The edge point set $\mathbf{X}$ and the transformed point set $\mathbf{Y}$. In both figures, the red "." indicates the points in the left image and the blue "." indicates the points in the right image.

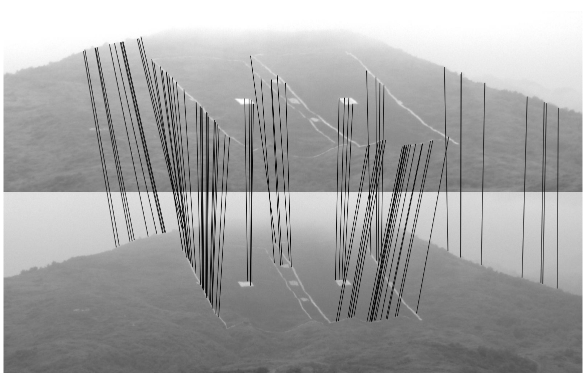

Figure 2. The proposed algorithm finds 86 correct matches

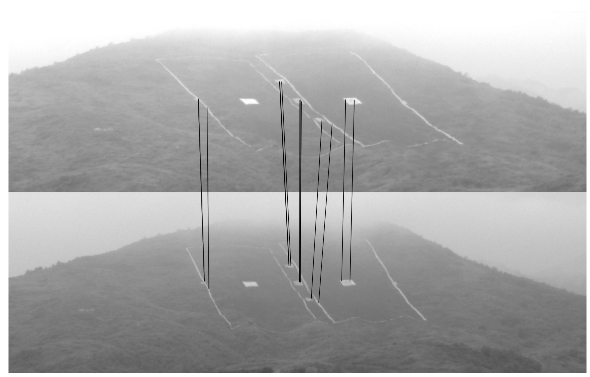

Figure 3. The ASIFT algorithm finds 14 correct matches

line according to the Eqs. (13) and (14) and evaluate the precision of the fundamental matrix estimation by the distances. Fig. (4) shows the distances.

$$d(\mathbf{u}_i, \mathbf{F}\mathbf{u}'_i) = \| \frac{\mathbf{u}_i^T \mathbf{l}_i}{\sqrt{l_1^2 + l_2^2}} \| = \| \frac{\mathbf{u}_i^T \mathbf{F}\mathbf{u}'_i}{\sqrt{l_1^2 + l_2^2}} \| \tag{13}$$

$$d(\mathbf{u}'_i, \mathbf{F}^T \mathbf{u}_i) = \| \frac{\mathbf{u}'^T_i \mathbf{l}'_i}{\sqrt{l'^2_1 + l'^2_2}} \| = \| \frac{\mathbf{u}'^T_i \mathbf{F}^T \mathbf{u}_i}{\sqrt{l'^2_1 + l'^2_2}} \| \tag{14}$$

In order to maintain the symmetry, we calculate the distances in two images at the same time. So in every experiment we get a single average distance through Equation (15).

$$\bar{d} = \frac{\sum_{i=1}^{n} [d(\mathbf{x}_i, \mathbf{F}\mathbf{u}'_i) + d(\mathbf{x}'_i, \mathbf{F}^T \mathbf{x}_i)]}{2n} \tag{15}$$

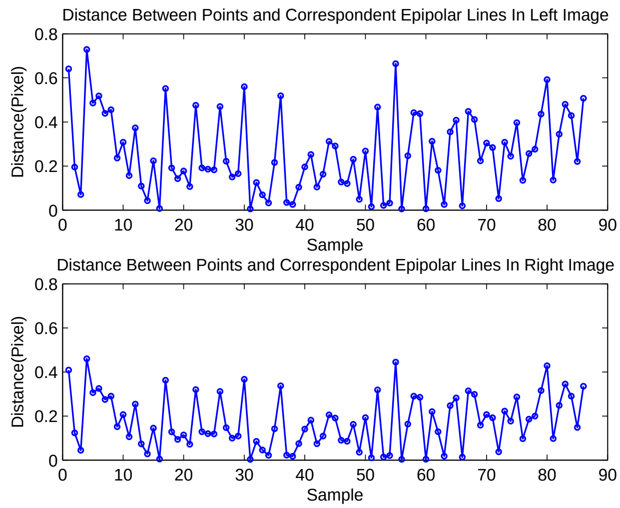

Figure 4. Results of a single experiment. Top: The distance between every point in Left Image and correspondent epipolar line according to Equation (13). Bottom: The distance between every point in Right Image and correspondent epipolar line according to Equation (14). The X coordinate of both figure means the point number in images.

We can then calculate the standard deviation of the distances according to Bessel formula.

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n}\{[d(\mathbf{x}_i, \mathbf{F}\mathbf{x}_i^{'}) - \bar{d}]^2 + [d(\mathbf{x}_i^{'}, \mathbf{F}^T\mathbf{x}_i) - \bar{d}]^2\}}{2n-1}} \tag{16}$$

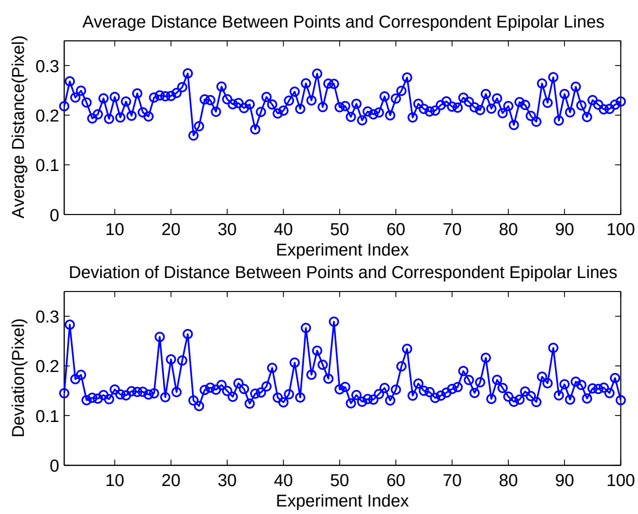

Figure 5. Results of repeated experiments . Top: The average distance between points and correspodent epipolar lines in every experiment according to Equation (15). Bottom: The deviation of the distance in each experiment according to the Equation (16). The X coordinate of both figure means the x-th experiment.

We repeat the experiment for 100 times, the average distance and standard deviation of every experiment are shown in Fig. 5. We can see that the average distances are all below 0.3 pixel and the standard deviations of the distance in every experiment are less than 0.3 pixel too.

4. CONCLUSION

We propose a method in this paper to estimate the fundamental matrix of a positing and monitoring binocular vision system with a long working distance and a large field of view. It is proposed to deal with the imaging blur introduced by the long working distance and bad weather, like fog, and absence of local features in the environment. Experiments indicate that our algorithm can work better than ASIFT under these special circumstances. We can then implement object matching, and self-calibration with the fundamental matrix.

REFERENCES

[1] Hartley, R. and Zisserman, A., [*Multiple view geometry in computer vision*], Cambridge university press (2003).

[2] Luong, Q.-T. and Faugeras, O. D., "The fundamental matrix: Theory, algorithms, and stability analysis," *International Journal of Computer Vision* **17**(1), 43–75 (1996).

[3] Fathy, M. E., Hussein, A. S., and Tolba, M. F., "Fundamental matrix estimation: A study of error criteria," *Pattern Recognition Letters* **32**(2), 383–391 (2011).

[4] Choi, S., Kim, T., and Yu, W., "Performance evaluation of ransac family," *Journal of Computer Vision* **24**(3), 271–300 (1997).

[5] Harris, C. and Stephens, M., "A combined corner and edge detector.," in [*Alvey vision conference*], **15**, 50, Manchester, UK (1988).

[6] Lowe, D. G., "Distinctive image features from scale-invariant keypoints," *International journal of computer vision* **60**(2), 91–110 (2004).

[7] Morel, J.-M. and Yu, G., "Asift: A new framework for fully affine invariant image comparison," *SIAM Journal on Imaging Sciences* **2**(2), 438–469 (2009).

[8] Matas, J., Chum, O., Urban, M., and Pajdla, T., "Robust wide-baseline stereo from maximally stable extremal regions," *Image and vision computing* **22**(10), 761–767 (2004).

[9] Denggao, L. and Kaihuai, Q., "Fast image registration algorithm based on randomized contour matching," *Journal of Tsinghua University(Science and Technology)* **46**(1), 111–114 (2006).

[10] Besl, P. J. and McKay, N. D., "Method for registration of 3-d shapes," in [*Robotics-DL tentative*], 586–606, International Society for Optics and Photonics (1992).

[11] Chui, H. and Rangarajan, A., "A feature registration framework using mixture models," in [*Mathematical Methods in Biomedical Image Analysis, 2000. Proceedings. IEEE Workshop on*], 190–197, IEEE (2000).

[12] Chui, H. and Rangarajan, A., "A new algorithm for non-rigid point matching," in [*Computer Vision and Pattern Recognition, 2000. Proceedings. IEEE Conference on*], **2**, 44–51, IEEE (2000).

[13] Jian, B. and Vemuri, B. C., "Robust point set registration using gaussian mixture models," *Pattern Analysis and Machine Intelligence, IEEE Transactions on* **33**(8), 1633–1645 (2011).

[14] Myronenko, A. and Song, X., "Point set registration: Coherent point drift," *Pattern Analysis and Machine Intelligence, IEEE Transactions on* **32**(12), 2262–2275 (2010).

[15] Ding, L. and Goshtasby, A., "On the canny edge detector," *Pattern Recognition* **34**(3), 721–725 (2001).

[16] Longuet-Higgins, H., "A computer algorithm for reconstructing a scene from two projections," *Readings in Computer Vision: Issues, Problems, Principles, and Paradigms, MA Fischler and O. Firschein, eds* , 61–62 (1987).

[17] Hartley, R. I., "In defense of the eight-point algorithm," *Pattern Analysis and Machine Intelligence, IEEE Transactions on* **19**(6), 580–593 (1997).